FW: results on CI for b

Brattin, Bill to: Hilbert, Timothy (hilbertj), Levin, Linda (levinls)   08/09/2012 06:25 AM

From: "Brattin, Bill" <brattin@srcinc.com>

To:

Cc: Benson.Bob@epamail.epa.gov, Berry.David@epamail.epa.gov

Tim and Linda

As indicated in the e-mail below, Leonid has determined that a fitting approach in which the b term is not constrained to be common across jobs may be better than assuming a common b.

Please take a look at his results and let me know what you think.

It seems that he calculated the 90% CO and asked whether it overlapped 0.4 (the best estimate of b in the common b approach).

Shouldn't he have considered the uncertainty in the common value of b as well, using a t-test or something similar?

Note: He just used an unweighted fit. Would things change in a variance weighted approach?

****

Bill Brattin

SRC, Inc.

999 18th Street Suite 1150

Denver CO 80202

Phone: 303-357-3121

Fax: 303-292-4755

e-mail: brattin@srcinc.com

From: Leonid Kopylev [mailto:Kopylev.Leonid@epamail.epa.gov]

Sent: Wednesday, August 08, 2012 9:02 PM

To: Brattin, Bill

Cc: Bob Benson; David Berry; Thomas Bateson; Krista Christensen

Subject: results on CI for b

Bill,

I got some time this evening, and since all of us here in NCEA will be away next week, I decided to get CI for b in R, so we could decide on next step without delay.

First of all, except Feeder, I got same or extremely close MLEs to your estimates. I couldn't get convergence for feeder as is, and had to multiply year by 10,000. As a result, I got a bit different estimates for a and b. Second, for 5 out of 7, confidence intervals lie away from 0.4, 3 being higher and 2 being lower. For one other (Expander) 0.4 is almost on the edge of confidence interval. Only for blender, confidence interval crosses

0.4, but then CI is rather wide, due to little data. (Technical note - I did 90% CI, but 95% CI would be a bit wider. Expander would be more consistent with 0.4, but rest would likely be quite similar.)

So, it seems to me that decay was not uniform across the tasks. Those with CI lower than 0.4 have much commonality, so a common lower value than 0.4 could describe these reasonably well.

However those with rapid decrease are quite heterogenous, so I am not sure about how to handle them.

Leonid

The data and my fitting results are in the spreadsheet I sent earlier
Here is a simplified version of the data.
The model for each job is: $C(j,t) = a(j)*exp(-b(j)*t)$ where t = (year of measurement – 1970)
I recommend you use the "adjusted conc" where non-detects are set to zero (highlighted in yellow)

****

Bill Brattin
SRC, Inc.
999 18th Street Suite 1150
Denver CO 80202
Phone: 303-357-3121
Fax: 303-292-4755
e-mail: brattin@srcinc.com

So, in excel, you don't get any estimate of variance - is this a problem?

I am on leave for the next 2 weeks starting Monday and I have my hands full. But could you send me the data and I'll try to run it before end of Friday.

The idea I had is to see if fixed b would be in confidence intervals for b-varied. If it were that way, I think we could live with fixed b having put forward that it
1) was consistent with UC assumptions
2) was statsitcially compatible with b fixed.

However, if for some process, b fixed were outside of CI for b -varied, that would be reason to use b-varied for that process or overall.

Leonid

"Brattin, Bill" ---08/08/2012 05:03:46 PM---No, the software I use (Solver in Excel) does not have a convenient way to do that. When that is nee

From: "Brattin, Bill" <brattin@srcinc.com>
To: Leonid Kopylev/DC/USEPA/US@EPA
Cc: Bob Benson/R8/USEPA/US@EPA
Date: 08/08/2012 05:03 PM
Subject: RE: Another Excel File for Discussion

No, the software I use (Solver in Excel) does not have a convenient way to do that.
When that is needed, I think it better to do it in SAS.
That could either be on your end, or I could ask UC to do it (although that comes out of their budget).
Let me know which you prefer.

****
Bill Brattin
SRC, Inc.
999 18th Street Suite 1150
Denver CO 80202
Phone: 303-357-3121
Fax: 303-292-4755
e-mail: brattin@srcinc.com

From: Leonid Kopylev [mailto:Kopylev.Leonid@epamail.epa.gov]
Sent: Wednesday, August 08, 2012 2:46 PM
To: Brattin, Bill
Subject: RE: Another Excel File for Discussion

Bill,

can you easily tabulate two-sided confidence interavls for a and b for fitted cases, confidence interval for a in fixed and note the value of b too?
I think they may be quite helpful in decision.

Thanks,
Leonid

From: "Brattin, Bill" <brattin@srcinc.com>

To: Bob Benson/R8/USEPA/US@EPA, Thomas Bateson/DC/USEPA/US@EPA, Danielle DeVoney/DC/USEPA/US@EPA, Leonid Kopylev/DC/USEPA/US@EPA

Cc: David Berry/R8/USEPA/US@EPA, Tim Hilbert <HILBERTJ@UCMAIL.UC.EDU>

Date: 08/08/2012 04:17 PM

Subject: RE: Another Excel File for Discussion

RfC Team

In accord with our discussions today, I have performed two supplemental analyses of the Marysville IH data, as summarized below.

Analysis 1: Fitting of Trionize data by job to exponential model, allowing slope (b) to vary between jobs.

Results are shown in attached file ("Trionizing Dept by job IH Samples 08012012 fit to exponential by job.xlsx"), on the second tab.
Although fits are achieved for all jobs, the upsweep gets pretty severe in some cases, and there would be a substantial difference between the fixed b and the job-specific b approach.
My instinct tells me fixed b is better, but I am not sure I could defend that position.

What do others think ?

Analysis 2: Are some IH values lower than reasonably expected?

I evaluated this by calculating what I think might be a reasonable lower limit for the analytical sensitivity S:

$$S = EFA / (FOVs * Afov * Q * d * 1000)$$

where
EFA = effective filter area (mm2)
FOVs = number of field of views counted
Afov = area of one field of view (mm2)
Q = sampling pump floe rate (L/min)
D = sampling duration (min)

I made the following guesses:
EFA = 385
FOVs = 100
Afov = 0.00785

Q = 10

I then "flagged" suspicious values where the reported concentration was lower than the "expected" S.
The results are shown in the attached file ("Conc vs sampling duration.xlsx"), second tab.
As seen, if my guesses are reasonable, then the occurrence of "flagged" results is fairly low...about 1-2% of the values.
All of these are samples reported as detects, and none are "non-detects"
Moreover, the magnitude of the difference between the "expected" and reported values is relatively small – an average of about 2 fold.

So, based on this, I conclude that even though some values do appear to be suspiciously low, they are sufficiently infrequent, and the difference is sufficiently small, that this is not likely to be a major cause for concern.

Opinions?

****

Bill Brattin
SRC, Inc.
999 18th Street Suite 1150
Denver CO 80202
Phone: 303-357-3121
Fax: 303-292-4755
e-mail: brattin@srcinc.com
[attachment "Trionizing Dept by job IH samples 08012012 fit to exponential by job.xlsx" deleted by Leonid Kopylev/DC/USEPA/US] [attachment "Conc vs sampling duration.xlsx" deleted by Leonid Kopylev/DC/USEPA/US] [attachment "data file for leonid.xlsx" deleted by Leonid Kopylev/DC/USEPA/US]

results for Bill on confidence intervals for b.xlsx